Can Acquiring Structured Data Improve a Learning Content Development Environment?

Wayne Gafford
Advanced Distributed Learning Co-Lab
Alexandria, VA
wayne.gafford@adlnet.gov

Schawn E. Thropp
Concurrent Technologies Corporation
Johnstown, PA
thropps@ctc.com

ABSTRACT

Structured data can be defined as a set of organized and identifiable pieces of information. The original underlying benefit of structured data focused on the separation of content from format. The separation enables content to be created independent of the delivery platform or rendering device. In its base form, structured data is represented by the Extensible Markup Language (XML) standard and is used in many specifications (e.g., S1000D, Darwin Information Typing Architecture – DITA, XHTML). The evolution of structured data has led to more benefits beyond the original separation of content from format concept. This evolution has set the foundation for data to exist in a beneficial ecosystem of data management practices.

The structured data benefits include improved metadata, linking, configuration, lifecycle management, content management and data exchange. However, for the Department of Defense (DoD), structured data must be acquired with requirements in mind for benefits to be realized. What must the learning, education and training (LET) community know about acquiring structured data? Can structured data improve a learning content management ecosystem? How can the LET community leverage the evolution of structured data practices to improve content management?

This paper discusses the benefits that can be achieved by acquiring structured data in the LET community. The paper concludes with guidelines for acquiring structured training data and the potential impacts on policy, standards and management.

ABOUT THE AUTHORS

Mr. Wayne Gafford combines a background in teaching and education with his experience in XML-based standards that has resulted in innovative ideas for e-learning content management and data interoperability. For the last five years, Mr. Gafford has led subcommittees, studies and projects that explore how learning content can benefit from structured markup and lifecycle support. Results have led to an increased awareness that standardized metadata and XML structure can unify diverse, but related content that support common systems, procedures and products. Mr. Gafford has taken his research to the Advanced Distributed Learning Initiative where he leads the Life Cycle and Configuration Team. He is the government co-chair of the S1000D Learning Standards Harmonization Task Team, is an active public speaker at S1000D and ADL events, and is a supporter of developing XML schemas that model instructional development and learning content to improve knowledge management and distributed learning. Mr. Gafford holds a Masters of Education from Marymount University.

Mr. Schawn Thropp is a Technology Advisor within the Learning and Human Performance Solution (L&HPS) division at Concurrent Technologies Corporation (*CTC*). He is responsible for research, development and advisement of advanced technologies within the learning, education, training and performance support domain. He has been involved with structured data, standards development and applicability within the learning and human performance domain for over 11 years. Mr. Thropp led the development and evolution of the ADL Initiative's Sharable Content Object Reference Model (SCORM®) along with the development of several international standards referenced within SCORM. Mr. Thropp has a Master in Computer Science degree from Johns Hopkins University.

Can Acquiring Structured Data Improve a Learning Content Development Environment?

Wayne Gafford
Advanced Distributed Learning Co-Lab
Alexandria, VA
wayne.gafford@adlnet.gov

Schawn E. Thropp
Concurrent Technologies Corporation
Johnstown, PA
thropps@ctc.com

INTRODUCTION

A 2009 Navy Inspector General (IG) Report on the state of computer-based training (CBT) found "minimal governance or standardization for the acquisition, design and development, or life cycle management of CBT curricula." The report concluded that "the necessary centralized governance and standardized management for courseware development, lifecycle management and content development have lagged behind the expansion of CBT." (Callahan, et al., 2009) The major improvement needed for technical DoD curriculum is a governance policy that calls for and defines a next generation concept of operations for aligning technical and training data for all acquired systems.

The relationship between governance, acquisition, design, development and lifecycle management of CBT, or any technical training content used in other environments must include a coherent and consistent approach to handling data on an enterprise level. Efficient business processes in a learning content management and development environment must be based on a quantitative set of expectations and requirements. Policy and processes that establish and maintain data governance will lead to front-end analysis processes, consistent implementation of training technologies, optimized lifecycle management practices and reductions in overall total ownership costs.

The use of structured technical content based on technical data specifications is one cornerstone to enabling best practices in a learning content management and development environment. This approach enables the mapping of metadata between acquisition, development and lifecycle management that results in learning content being **ensured**, **traceable** and **aligned** to the systems and to the learners. Evidence from a Naval Education Training Command (NETC)-funded report, *Computer-Based Training & Personal Computer-Simulation Prioritization and Cost Estimation Assessment for NETC COO Supporting FY10/11 Spend Plan and POM12 Submittal*, suggests that there is quantitative evidence that integrated structured markup is warranted. The report stated, "Of the 408 projects submitted for Computer-Based Training and PC-Simulation Maintenance, two-thirds reported the primary reason for the maintenance request was due to equipment or publication changes. Better integration of technical information with training would alleviate some of this rework." (NETC, 2009)

Structured content defined by the right community of practice-based data specification is the definitive tool to harness value-added measures of government data rights and data management.

PROBLEM STATEMENT

Technical learning content focused on operation, maintenance and troubleshooting of equipment in the DoD is not consistently aligned to authoritative source information in regards to format, content management or metadata. The misalignment begins at acquisition when technical data and training data requirements are not developed in synchronized processes. Inevitably, the learning content development environment falls into a perpetual state of delayed production schedules disconnected from the body of materials that inform training requirements.

Can acquiring structured data improve a learning content development environment and address the governance findings in the 2009 Navy IG report?

This paper examines:

- the attributes and benefits of structured data as critical evidence for supporting the conclusion, and

- the association of structured data practices to data acquisition activities

STRUCTURED DATA

Structured data, or structured markup, as it is known today, began in 1969 with IBM's creation of GML – understood to stand for Generalized Markup Language (Kay, 2005). GML supported text editing, formatting and information-retrieval systems that needed to share data. Several years later GML was expanded to support additional concepts, such as short references, link processes and concurrent document types, and was released as an International Organization for Standardization (ISO) standard - ISO 8879:1986 Standard Generalized Markup Language (SGML). SGML was designed for information processing by separating the content of a document from the format of the document by creating a formal representation of data within a file.

The critical aspect of structured markup is the rules to govern how the content is physically structured in a valid document. The content structure rules are always based on a unique set of community-of-interest specifications. The important point to SGML as a standard is it sets rules *for creating rules about unique document structures.* This is why SGML is applicable to any industry that wants to govern its unique set of content. SGML specifications have been developed for the military, the Library of Congress, the American Chemical Society and other industries to improve information processing and management.

Structured Markup Breakthroughs: HTML and XML

The breakthrough SGML-based specification was the HyperText Markup Language (HTML) first introduced in 1991 by Tim Berners-Lee. HTML grew into the standard markup for content viewed in a web browser. As demand grew for viewing and processing data over the web, SGML proved to be too large of a standard designed primarily for documents. The eXtensible Markup Language (XML) was written as a simplified version of SGML for usability on the web. XML's simplicity broadened the applicability of structured markup beyond documentation and into information processing, data exchange and web services. A variety of structured data specifications exists today. **Table 1** lists a set of common structured data specifications.

Structured markup is the key tool that has enabled the world to see and exchange content of any kind for any purpose. There are hundreds of XML applications that structure content and serve as the lifeblood of internet-based economies and social networks around the world. Most important, structured markup used for documentation and information processing on the web is substantiated through a governance model in the form of the World Wide Web Consortium (W3C). The W3C is a testament to structured markup's longevity, growth and future directions. For the DoD to take control of learning content acquisition, design, development and lifecycle management, it should re-evaluate its own governance, policies and practices on the use of XML.

Table 1. Common Structured Data Specifications

Structured Data Example	Description
HyperText Markup Language (HTML)	HTML is the publishing language of the world wide web (Raggett, 1999). It is structured data used from a publication perspective.
S1000D	S1000D is an international specification for the production of technical publications utilizing a common source database. (S1000D, n.d.)
Darwin Information Typing Architecture (DITA)	DITA defines a set of document types for authoring and organizing topic-oriented information, as well as a set of mechanisms for combining, extending and constraining document types. (DITA, n.d.).

THE BENEFITS OF STRUCTURED CONTENT

Standardized management for content development and lifecycle processes begins with structured markup that reflects attributes of the content and the business processes of the enterprise. It can establish a predictable pattern of file naming conventions, unique identification of configurable assets, business rule implementation and shared enterprise content management infrastructure. Learning content developed in an XML rule-based environment does not inhibit creative instructional design, multimedia development and tailored assessments. These types of learning content can be written for and utilized in Computer Based Training, SCORM, mobile platforms, instructor-led training (ILT), tactical training and job performance aids.

As structured content supports tailored learning content, another ingredient to the learning development environment is necessary: the common source database (CSDB). The CSDB concept collects all related content into a centralized point of storage. Data inventory control and reuse are assured across information types and security is improved. Structured markup and CSDBs are the two primary building blocks to improving a learning content development environment when lifecycle and configuration management are required.

Separation of Content from Format

Learning content development environments require flexibility. Structured markup enables flexibility because the content is not trapped by a format that is not designed for a specific community of practice. A proprietary format designed for general purpose use cannot offer the data typing an industry-based structured markup specification can offer. Format is supplemented with context in structured markup and offers freedom of choice for data development and processing. Choices include:

1. **Device Rendering:** Structured markup allows content to be deployed to different devices without locking the content into the device. The integrity of the data is maintained while adding value through platform neutrality.
2. **Style Rendering:** Structured markup allows for a diversity of styles to be applied to the same data. This benefit is closely related to device rendering as each platform (mobile, web, paper) would each require its own content style. The structured markup in the CSDB ensures a single data entity can be processed through tailored styles.
3. **Write Once:** Structured markup allows content to be isolated as a single entity for multiple purposes. This benefit is the principle of minimal development of content for maximum usage across context and styles. It is also the economic principle guiding the reduction in total ownership costs.
4. **Multi Lingual:** Structured markup allows for effective transformation of content from one language to another. A single-source language may serve as the primary language – setting the foundation for a multi-lingual CSDB. Structured markup can support the translation component and as the output for the target language.
5. **Description:** Structured markup describes the meaning of content, not the format of the content. The semantics in the structured markup adds business value to the data and enables machine-readable processing. State-based information set into markup can determine the content sequencing for interactive user experiences.

Structured Interactivity

One major appeal of web-based information is quick access to related content. Related information may be drawn together to enrich a learning experience. **Table 2** describes the benefits of linking and content aggregation for learning content development environments.

Table 2. Linking and Aggregation

Relationship Type	Benefit
Linking	Linking content from one file to another is achieved through structured markup. It is a form of processing that does not rely on transformation, rendering or devices. It serves as an enhancement to the learning experience by offering quick jumps from file to file using structured markup.
Aggregation	Structured markup can serve to assemble and reassemble content. In the case of aggregation, the markup does structure content within a document but arranges the files into a block of documents then sequences the viewing of the content. The structured markup consists of references to files that are called upon according to the design of the instruction.

Data Lifecycle Management

Critical to DoD learning content development environments is the ability to track content as changes to systems are made over time. Much of the learning content in DoD is technical in nature. The content would not exist unless a system is in place to deploy and use it. The technical publication/manual community is equipped to handle traceability and synchronization of technical data to weapon systems and has broad structured markup practices in place to ensure data readiness.

The implication to data lifecycle management is Warfighter readiness. Learning content is often out of alignment with fielded systems. As noted by the Navy IG's report, "Time to affect a change in CBT curricula can be protracted; eighteen months is not unrealistic as an average time from the point a need for courseware correction is identified to the time the revised product is contracted, developed and delivered." (Callahan et al., 2009) Structured markup for technical learning information stored in a CSDB shared with related technical publications can improve learning content development environments and Warfighter readiness.

Lifecycle Management Stages

Structured learning content can serve each stage of systems acquisition and data lifecycle management. During the *design phase*, manpower, personnel and training needs analysis content can be structured and mapped to fault tree analysis, projected maintenance requirements and system safety procedures. This mapping *ensures* links between system design data and training requirements. During the *production phase*, early curricula can be structured into markup to create product support packages to include shared data from technical publications. This mapping creates *traceability* between design phase output and product support packages. During the *testing phase*, drafted content can be rapidly accessed and resynchronized as system test results are reported. Traceability provided by structured markup is a key element. During the *support phase*, deployed learning content directly impacted by a design change can keep pace with new system configurations using metadata as a real-time discovery tool. This mapping creates long-term *alignment* between fielded systems and product support packages. These relationships are discussed further in the "Structured Markup and Data Acquisition Policies" section of this paper.

To realize these benefits, the government must acquire learning content using statements of work (SOW) that detail acquisition requirements. Naming a specification, identifying the use of optional metadata in the specification for use in automated business practices, establishing a threshold for data reuse, and requiring the use of a government-owned CSDB must be articulated in the SOW, especially if the government intends to own the data.

The Link to Technical Data Rights

Structured markup is an enabling technology that allows technical training content to effectively interoperate with related technical data across system lifecycles. The use of a markup specification is also critical to technical data rights. If the DoD is to own technical data packages that support purchased systems, it must proceed with a coherent acquisition strategy where data management and development activities are consistently applied across the enterprise. XML tidies up an unwieldy DoD enterprise proposition for managing a diverse array of data in a diverse array of formats stored in disconnected environments. This is a tall challenge on a broad scale but not insurmountable at a program level. There is evidence that programs and industry are willing to support DoD policies that require structured markup, especially when those policies foster cooperation and integration between the technical data and training communities. The evidence resides in a survey sponsored by the Office of the Secretary of Defense (OSD) for Personnel & Readiness and conducted by the Naval Postgraduate School (NPS) from October 2009 – May 2010. It focused on the use of an XML-based data markup specification, S1000D. S1000D is designed to support technical and learning data for air, land and sea systems. NPS asked DoD programs and industry, "Should S1000D be required by the Department of Defense?" (Blaise et al., 2010)

NPS SURVEY ON S1000D STRUCTURED MARKUP

NPS structured the survey into two formats: face to face (F2F) interviews and online questionnaires. The F2F interviews included 20 interviews with 24 interviewees. The online questionnaire added another 180 participants. Interviewees were from organizations across all four military services, the Coast Guard, defense industry, S1000D product vendors and S1000D governance bodies. The survey participants represented experienced, practical and informed judgments of structured data users. The

Navy/Marine Corps constituted the largest group, representing 43% of the total, followed by "Other" at 27%, the Air Force with 21%, and the Army and Coast Guard with approximately 5% each.

A full representation of the survey analysis cannot be given justice in this paper, only the final high level results and the commentary made on the use of S1000D markup for integrated data. The main question, "Should S1000D be required by the DoD?", was geared toward technical publication communities; however the survey documented the reported benefits to the training community. Final analysis concluded with the following data:

Question: Should the DoD require S1000D for the following contexts? Percentages indicate agreement with the three levels of requirements:

- All DoD systems? (40.6%)
- New acquisitions but not legacy? (62.7%)
- Large CAT (I/II) but not small acquisitions? (51.3%)

Of these respondents, 74% believed S1000D markup provides a benefit for linking logistical/technical data and SCORM. Although not all DoD training content is published into SCORM, the informed judgment that structured markup can forge connections between technical data and training content is noted.

Learning Content Structured Markup and SCORM

Learning content benefits from the additional rigor of XML implementation beyond that which is already instituted through SCORM. SCORM does not apply any structured markup requirements for training content. SCORM is content and format agnostic. This design is on purpose. SCORM is essentially a content packaging specification with support for learner tracking and sequencing in a learning management system (LMS). Applying structured markup to content fills a gap intentionally left in SCORM that actually results in dramatic lifecycle issues. These issues are observed in the 2009 Navy IG on the state of CBT and can be addressed by the use of S1000D markup as a basis for enterprise learning content management strategies.

Survey Recommendations

NPS concluded from the 200+ respondents that DoD should require the use of S1000D markup. The conclusion included a caveat that takes the benefit of structured markup and elevates it into an enterprise data management vision driven by policy. NPS recommended "the establishment of a DoD Technical Information Governance Office to provide top-level, enterprise-wide leadership and guidance for technical publications and technical data across acquisition, logistics, maintenance, training, and other relevant endeavors." (Blaise et al., 2010)

It is important, therefore, to baseline the prominent DoD data policies on structured markup for content in general and articulate the specific directions for markup on learning content. The mission of the Governance Office could encompass helping to fill the gaps between the baselined policy guidelines on structured markup and the desired acquisition best practices for learning content regarding improved management and development environments.. Practically speaking, there must be a reduction in total ownership costs for a markup strategy to be attractive.

STRUCTURED MARKUP AND DATA ACQUISITION POLICIES

There are three principal DoD policy documents that provide guidance on structured markup. Guidance ranges from the generic, to the implied and to the specific. Taken as a whole, the policy language signals to industry and to DoD programs that XML structured markup is a desired data management strategy without explicitly requiring the use of a particular specification. The policy documents are:

- DoD Instruction 5000.2 - Operation of the Defense Acquisition System
- The Defense Acquisition Guidebook
- DoD Directive 8320.02 - Data Sharing in Net-Centric DoD

Below is a look at what each document says about "data management" and "training" with respect to the benefits of structured markup.

DoD Instruction 5000.2 - Operation of the Defense Acquisition System

DoD Instruction 5000.2 divides the DoD systems acquisition process into three milestones (MS): A, B and C. These milestones are subdivided into activities built upon outputs from previous phases. The phases are:

- MS A
  - Pre-Milestone A: Materiel Solution Analysis
  - MS A: Technology Development

- MS B
  - Engineering and Manufacturing Development
- MS C
  - Production and Deployment
  - Operations and Support

The highest, most generic level of direction for data management and training is cited in DoD Instruction 5000.2. In section 5, para 7(g), it is simply stated that a technical development strategy will include a data management strategy in MS B. (OSD AT&L, 2008)

Section 8(c)(1b) directs that lifecycle sustainment considerations in MS C include data management. (OSD AT&L, 2008)

Enclosure 12, section 9, Data Management and Technical Data Rights states in para (a), "Program Managers for ACAT I and II programs, regardless of planned sustainment approach, shall assess the long-term technical data needs of their systems and reflect that assessment in a Data Management Strategy (DMS)." (OSD AT&L, 2008)

5000.2 references "training" at a higher level. In Enclosure 2 – Procedures, within Section 8 - Operations and Support Phase C – Phase Description, (1) Life Cycle Sustainment, Optimize Operational Readiness, para (a) only says:

"(a) Human-factors engineering to design systems that require minimal manpower; provide effective training; can be operated and maintained by users." (OSD AT&L, 2008)

Connecting MS Activities to S1000D Markup

References to markup are not included in 5000.2. The 5000.2 serves more as an acquisition framework leaving the implementation details to other guidance and implementation policies that focus on a narrow slice acquisition. However, it is important to draw relationships between milestone activities and S1000D markup. Standards must have relevant business processes for any reasonable benefit to be manufactured.

Milestone A – Materiel Solutions Analysis and Technology Development Phase

Activities in this phase include creating functional system definitions, understanding failure modes, developing reliability centered maintenance, human systems engineering and lifecycle sustainment planning. The bridge from these activities to training is front-end analysis for human performance requirements. Structured markup does not have a direct effect on front-end analysis as the output is based on factors outside of file formats, such as system performance requirements. In this case, the markup can capture the front-end training analysis based on and linked to MS outputs. If done coherently, the front-end training analysis can be linked to the resulting structured curricula and performance support. This is the start of a true data ecosystem in the training space.

How do MS A activities relate to S1000D markup? Through *ensurance.* We can *ensure* that logistical support analysis outputs from MS A resulting in technical documentation and training can be ensured, managed and protected by S1000D markup. Markup is the root for consistent implementations of training technologies.

Milestone B – Engineering & Manufacturing Development Phase

Activities in this phase include the development of product support plans, running critical design reviews, development and operational testing, and verifying initial product baselines. The link from these activities to training is the coordinated and gradual development of technical and training content in S1000D markup as the system takes shape. The manpower, personnel and training needs analyses can be crafted into curricula, learning modules, performance support and other product support data.

Central to MS B is change. System designs are put to the test during engineering and manufacturing. Baselines evolve into new versions as reviews and tests are completed. Documentation must be created in the form of test reports, early product support elements and procedural steps.

How do MS B activities relate to S1000D markup? Through *traceability.* Documentation keeps up with developmental testing enabling agile *traceability* between evolving systems and product support elements configured by XML markup. Structured markup is the lynchpin that allows product data developers to grow a data ecosystem in an enterprise environment.

Milestone C – Production & Deployment and Operations & Support Phase

Activities in this phase include low rate of initial production (LRIP), joint interoperability testing, analysis of system deficiencies to determine corrective actions, full rate of production and product support adjustments. The link from these activities to training is the sustained alignment between product and product support.

In many DoD cases, technical manuals are updated after engineering changes have been approved. The validated and verified technical data is published and distributed to a supply system. Learning centers are often placed on technical manual distribution lists. The result is the determination of training requirements after the system is fielded. Curricula updates are protracted and readiness is lost.

How does S1000D markup relate to the MS C systems acquisition processes? Through *alignment.* Structured markup provides the ability to make baseline *alignments* between the configuration of the technical data products to the baseline configuration of the product design at the point of manufacturing. Structured Markup is the lifeblood between fielded systems and sustained product support.

The Defense Acquisition Guidebook (DAG)

The DAG provides a more focused and pronounced message about data management and training than what the 5000.2 understandably offers. The Defense Acquisition Guidebook is designed to complement 5000.2 by providing the acquisition workforce with discretionary best practices that should be tailored to the needs of each program. (DAU, 2010)

The central data theme in the DAG is fostering integrated data environments (IDE). Each reference to data management includes training content in the broadest sense. The continuity of each reference with respect to structured markup is punctuated by a specific mention of structured markup and S1000D.

The references to data management include the following sections:

2.2.14 – Data Management Strategy (DMS) and Technical Data Rights

> "DMS should reflect the assessment and integration of the data requirements ***across all the functional disciplines***…" (DAU, 2010)

> **2.3.14.2 – Integrated Data Environment**

> "Program Managers should establish a data management system within the IDE ***that allows every activity involved*** with the program to cost-effectively create, store, access, manipulate, and exchange digital data." (DAU, 2010)

> **5.1.2.1 – Key Program Documents**

> Acquisition Strategy – "…it should address how the ***product support package*** required to support the materiel management, distribution, technical data management, support equipment, maintenance, ***training***, configuration management, engineering support, supply support, and failure reporting/analysis, functions will be acquired." (DAU, 2010)

References to training in the DAG are oriented more to MS C lifecycle sustainment activities and do not refer to MS A and B activities. However, the main thrust of the training references is rapid synchronization between updated product baselines and updated product support packages. The references include:

> **5.4.2.2.1. Initial Life-Cycle Sustainment Plan**

> Supportability Analysis Process – "Training and HSI requirements, including the ***training requirements/objectives (for both operator and maintenance training) relative to training courses, materials, and training equipment*** to enable personnel to effectively perform tasks supporting the Concept of Operations (CONOPS) and the maintenance concept. The requirements for specific training strategies to be used to meet the Sustainment Key Performance Parameter (KPP), such as distance learning should also be addressed." (DAU, 2010)

> **5.4.2.2.2. Maintenance & Sustainment Strategy Development**

> "…enablers can range from system design features (e.g., condition based maintenance) to supply chain features (e.g., ***rapid distribution of***

> ***tailored support packages, just in time training / distance support***, total asset visibility anywhere in the support chain, dedicated rapid response support teams ***analyzing real time data***." (DAU, 2010)

Although the training references are not couched in a context about management and development, the markup and specification issue is implied through content sustainment planning. The assertion provides a literal interpretation of the DAG's specific reference to S1000D and to 8320.02, Data Sharing in a Net-Centric DoD's call for "semantic tagging". The DAG refers to structured markup and S1000D in the following section:

> **4.2.3.1.7.1. Data Acquisition**
>
> "Data acquisition encompasses all activities that create, obtain, or access data from internal or external sources to satisfy data requirements driven by the data strategy. When at all possible, data should be acquired in a ***structured format that is independent of the method of access or delivery and defined by or based on open standards.***" (DAU, 2010)
>
> Consider the following standards for defining the structure of digital data:
>
> ***S1000D International Specification for Technical Publications Utilizing a Common Source Database (http://www.s1000d.org/) .*** (DAU, 2010)

The DAG specifies the need for an IDE across functional disciplines and clearly implies that training is a firm part of technical data packages. Structured formats independent of access delivery is a key principle. It can therefore be asserted that the DAG recommends S1000D as a digital data format for the improvement of IDEs, including learning content environments.

DoD Directive 8320.02 – Data Sharing in a Net-Centric DoD

The DAG suggests that "compliance with the DoD Directive 8320.02, "Data Sharing in Net-Centric DoD" and the DoD Net-Centric Data Strategy is an essential prerequisite of net-centric operations." (DAU, 2010) The purpose of the 8320.02 is to "direct the use of resources to **implement data sharing** among information capabilities, services, processes, and personnel interconnected within the Global Information Grid (GIG)." (ASD NII/DoD CIO, 2007)

Structured markup is a key ingredient to a strategy that unifies common content in an IDE in cost-effective ways. The 8320.02 uses different vernacular to describe structured markup. The Directive sets data policy in section 4, para 4.2 by stating: "Data assets shall be made visible by ***creating and associating metadata ("tagging")***, including discovery metadata, for each asset." (ASD NII/DoD CIO, 2007)

The 8320.02 takes the concept of structured markup a step further by realizing technological benefits do not happen without human action. Improvements to learning development environments are conscious decisions born out of recognition of problem statements that impede missions and reduce readiness. It is policy as stated in para 4.7 that "semantic and structural agreements for data sharing ***shall be promoted through communities*** (e.g., communities of interest (COIs)), consisting of data users (producers and consumers) and system developers." (ASD NII/DoD CIO, 2007)

DoD Directive 8320.02 connects structured markup (semantic tagging) with the direction to promote its use. However, the true promise for integrated data including learning content using structured markup must pass this key litmus test before full acceptance is given: Does structured markup reduce total ownership costs (RTOC) for DoD?

STRUCTURED MARKUP COST BENEFIT ANALYSIS (CBA)

In 2009, Advanced Distributed Learning collaborated with the Institute for Defense Analyses (IDA) to perform a cost benefit analysis on the use of S1000D for integrating technical learning content and technical data in a common source database. The study is part of the larger "Bridge Project" funded by the AT&L/RTOC program. The primary goal of the project is to improve acquisition, configuration and lifecycle management of technical learning content.

The study considered four primary technical principles developed in the Bridge Project against a single metric: reduction in the number of hours required to develop and manage technical information. The factors are:

- All program technical information structured in S1000D

- All program technical information stored in a common source database
- Use of the Bridge API to connect learning content development tools with common source databases
- Use of a query tool to identify all reusable technical information needing review according to system design changes.

Data Samples

IDA conducted separate cost-benefit analyses from two perspectives: OSD and individual program. The OSD perspective recognizes OSD's broad interest in seeing whether the new Bridge business model will lead to net cost savings—benefits exceeding costs—if implemented by the Navy and other Services as a whole. Analysis of this perspective was conducted for an "aggregate" sample: the Navy's yearly production of all Hull, Mechanical, and Electrical (HM&E) technical manuals produced by the Naval Ship System Engineering Station (NAVSSES) in Philadelphia, and all Computer-Based Training (CBT) courses delivered by Navy eLearning (NeL), a part of the Naval Education and Training Command (NETC).

The second perspective reflects the focus of Program Offices on their individual systems of interest. It focused on the benefits of a "single-system" sample: technical publication and training content for the AN/AQS-20A mine-hunting sonar for the Littoral Combat Ship.

Methodology

The cost savings for producing future technical manuals and training courses using the Bridge business model were calculated by estimating the savings for producing a nominal technical manual and a nominal training course and then scaling up the results to the production of manuals and courses for the aggregate and single-system samples. (Levine, 2010)

The Submarine Learning Center in Groton, CT provided a list of 80 steps required to develop one hour of curriculum. NAVSSES provided a list of steps required to develop a 500-page technical manual. The steps were compared to a "to-be" integrated data development and management process based on structured content and the other Bridge principles. Staff hour estimates needed to perform the detailed tasks for the nominal products were captured. Estimated staff hours needed to perform content development and management according to the Bridge business model were subtracted from the as-is system. The savings were calculated by applying average hourly pay rates for technical writers and course developers.

Analysis Results

The IDA study is the first known quantitative analysis on the financial benefits of structured markup on learning content development environments. The primary benefits for both technical publications and learning environments were derived from response times to engineering changes and the implementation of reusable data. The study estimated a reduction of 490 technical manual lifecycle support hours, or 6%. To support one hour of courseware, the study estimated a reduction of 66 lifecycle support hours, or 16%. (Levine, 2010)

The analysis suggests that the Navy could achieve savings by integrating the future production of technical manuals and training courses using S1000D markup and the other Bridge factors. The estimated 10-year savings (benefits less costs) ranged from $32.0 million to $165.5 million for integrating the yearly production of all HM&E technical manuals and CBT courses delivered by NeL. (Levine, 2010)

The single-system analysis suggests that applying S1000D markup and the other Bridge factors to the AN/AQS-20A mine-hunting sonar for the LCS would produce 10-year benefits of $307,700. This is the net savings if the full investment and implementation costs were already paid. If not, it would take 5.8 years to cover the $1.8 million cost of investment, plus some additional years to cover the small but uncalculated implementation costs for the AN/AQS-20A alone. (Levine, 2010)

Analysis Results Implications

The use of structured markup in the learning development environment will have a creative impact on the composition of development teams. Structured markup would be factored into the instructional systems design (ISD) and development processes where reuse, linking and interactivity is concerned. Instructional designers normally assigned for courseware development will need to be augmented by programmatic support that can translate their work into compliant formats.

By using structured markup throughout their content development environment, programmers will add permanent efficiencies and value to key business processes. For example, SCO Workbench (http://www.openscorm.org/) is an open source training development tool that uses structured markup behind a user interface. The Bridge Project is supplementing SCO Workbench with S1000D course structure and markup support. The Bridge Project is doing the same with the Navy's Authoring Instructional Materials (AIM) development software.

The strategic approach for use of structured markup in ISD environments is to shield the instructional designer from the markup and allow design and production work in an interface. SCO Workbench and AIM capture data out of specific fields and pull-down menu options then insert them into structured markup behind the scenes. The structured markup in this case becomes the asset to a coherent data management strategy that treats the data as business assets in coordination with related information. Complexity is reduced through planned automated business process. If there is a desire to integrate with a pre-existing courseware system of any kind, the structured markup actually helps eliminate the data interoperability problem through the use of an API that counts on the expected rules and formality that structured markup provides.

CONCLUSION

Benefits to acquiring structured markup in a learning content development environment are signaled by DoD data policy, voiced by respondents to an NPS survey on the use of S1000D and quantified by an IDA CBA. Two out of three changes to curricula in the Navy are driven by equipment changes. It can be inferred that there is a similar percentage in the other services, although specific analyses are unknown.

The single largest cost for training content and technical manuals is lifecycle maintenance. Structured markup reduces ownership costs through the ability to quickly discover the training to review traced to proposed system design changes. Training content, stored in the same class of structured markup aligned to other related information products, will be maintained concurrently and not after new system updates are fielded. Reduced ownership costs become linked to readiness levels when data acquisition strategies are based on XML specifications designed for a particular community of practice. Government must fund sustainment activities for structured markup to be a benefit to learning content development environments. Structured markup is a pivotal tool for centralized governance and standardized management of DoD courseware development and lifecycle management.

REFERENCES

ASD NII/DoD CIO, (April 2007). "*8320.02 Guidance for Implementing Net-Centric Data Sharing*"

Blaise C., Falby J., McCauley M. (Aug 2010), *Should S1000D be Required by the Department of Defense?*

Callahan, CAPT K., et.al. (March 2009). *NAVINSGEN Computer Based Training Study*.

DAU. (August 2010). "*Defense Acquisition Guidebook*".

DITA. (n.d.). Retrieved June 10, 2011, from http://docs.oasis-open.org/dita/v1.2/os/spec/introduction/dita-1.2_complete_overview.html

Kay, R., (August 8, 2005). Quickstudy: Markup Languages, Retrieved from http://www.computerworld.com/s/article/103711/Markup_Languages?taxonomyId=11&pageNumber=1.

NETC Computer-Based Training & Personal Computer-Simulation Cross-Functional Team, (August 2009). *Computer-Based Training & Personal Computer-Simulation Prioritization and Cost Estimation Assessment for NETC COO Supporting FY10/11 Spend Plan and POM12 Submittal*, Pg 8.

OSD/AT&L, (December 2008). "*DoD Instruction 5000.2, Operation of the Defense Acquisition System*".

Raggett, D., Le Hors, A. & Jacobs, I. (December 24, 1999). HTML 4.01 Specification, Retrieved June 10, 2011, from http://www.w3.org/TR/html401/

S1000D. (n.d.). Retrieved June 10, 2011, from http://public.s1000d.org/Pages/Home.aspx

Standard Generalized Markup Language. (n.d.). Retrieved June 10, 2011, from http://en.wikipedia.org/wiki/SGML

W3C. (n.d.). XML Essentials. Retrieved June 10, 2011, from http://www.w3.org/standards/xml/core